A Stereo-vision based Object Tracking Approach at Roundabouts

Maximilian Muffert, David Pfeiffer, and Uwe Franke

***Abstract*— This article presents a stereo-vision based system for the recognition of dangerous situations at roundabouts. To this end, we investigate the necessary field of view and viewing direction using videos taken by a panoramic camera. Using the insights of these tests, we build up a stereo-vision system. This system is based on the well established disparity estimation scheme semi-global matching and the recently introduced medium-level representation called Dynamic Stixel-World. This data is the input to compute a time-to-contact measure that makes explicit use of the roundabouts structural characteristics. This measure enables us to create a system for driver warning or possible automated intervention. Our empirical studies reveal that the warning decision correctly mimics human driver decisions in most roundabout scenarios.**

I. INTRODUCTION

The number of cars in inner city traffic has increased steadily over the last 20 years. For optimizing the traffic throughput in inner city environments, the use of roundabouts (see Fig. 1) has proven as a successful method. Especially in Europe, this strategy of traffic management is very popular. For example, there are over 20.000 roundabouts in France. At roundabouts the number of hazard points is significantly smaller than on classical intersections without traffic lights[1] (see Fig. 2). Further theoretical advantages are described in detail in [11].

Nevertheless , dangerous situations for drivers could still occur. Assuming right lane driving, the following points summarize the major risks of a driver entering a roundabout:

- Overlooking traffic from the left side.
- Misjudging the speed of oncoming cars.
- Misjudging the driving behavior of oncoming cars.

Furthermore, it is not advisable to rely on the turn signal of oncoming cars. A study of the European Car Club has shown that more than one out of three drivers did not use the turn signal correctly in roundabout scenarios[2].

The two illustrated situations in Fig. 1 underline the named problems for a waiting driver: It is challenging to predict if an oncoming car will leave or will pass the ego-vehicle. Only in the last time-step $t = 2$ it is obvious that the left vehicle turned in contrast to the right vehicle which passed the ego-vehicle position.

In view of this challenge, it is our goal to develop an advanced collision avoidance systems (ACAS) that helps to reduce the risk of accidents in critical roundabout scenarios as seen in Fig. 1.

[1]http://www.nuernberg.de/internet/verkehrsplanung/z_konfliktpunkte_kreisverkehr_0_8.html, viewed: 30.09.2012

[2]http://www.ace-online.de/der-club/news/schlechte-blinkmoral-auf-deutschen-strassen/newssearch/blinkmuffel.html?cHash=ddb6d4c28a, viewed: 30.09.2012

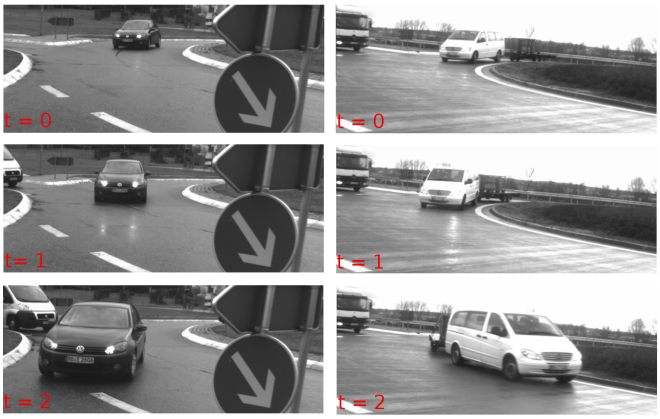


Fig. 1: Two typical situations at a roundabout. The left image sequence shows a turning vehicle. In the right sequence an oncoming car passed the ego vehicle position. At the first two time-steps $t = 0$ and $t = 1$ the driving intension of the oncoming cars is not obvious.

The development of ACAS is one of the most active and most challenging fields of research in the automotive community. In this context, the detection and tracking of other traffic participants is of particular interest. Next to radar [16], [21] and lidar based methods [24], [14], stereo-vision [2], [7], [8], [13] has already proven to be a powerful solution for this task. In Erbs, et al. [7] a multi-class traffic scene segmentation approach is described which estimates the pose and the velocity of moving objects like vehicles, pedestrians or bicycles from a mobile platform. In addition Barth, et al. [2] describes an approach which is based on tracked 3D feature points to extract the complete motion state (pose, velocity, acceleration and yaw rate) of vehicles.

Next to the detection and tracking, it is inevitable to estimate the potential collision risk between the ego vehicle and oncoming cars for ACAS. Berthelot et al. [4] present a stochastic approach to estimate collision probabilities as well as time-to-contact (TTC) probabilities to handle complex traffic situations. It is assumed, that the motion state (and its accuracy) of involved vehicles is known. The approach is evaluated in turn-left and crossing scenarios. In [22], a collision prediction for intersections is described which is based on the analysis of traffic video data. Learned traffic patterns are used for collision computation and for the detection of traffic conflicts.

In this article a Stop-and-Go Assistance System for the special case of roundabout scenarios is described. It is an extended version of [17] and includes tracking and detection

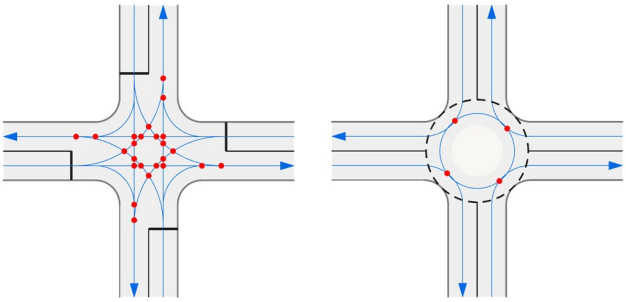


Fig. 2: The number of conflict nodes at normal 4-way-intersections (left) and at roundabouts (right): In contrast to intersections (24 critical points), roundabouts have only 4 critical points.

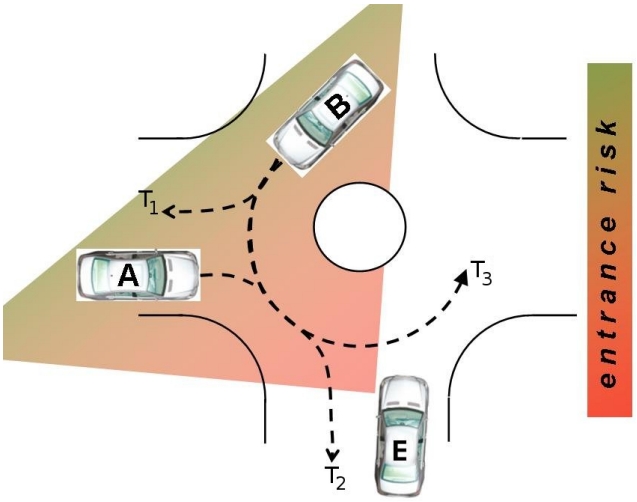


Fig. 3: The entrance risk at roundabouts grows with descending distance. The ego-vehicle E has to consider that both cars (A and B) will pass. Altogether, this is a very challenging situation for the environment perception.

schemes as well as the computation of the potential collision risk between the ego vehicle and oncoming cars which is expressed by the TTC.

In contrast to trajectory and collision prediction at intersections [22], roundabouts exhibit many different complex (non-straight) driving routes that other traffic participants can take (see Fig. 3). Unfortunately, this causes problems for today's ACAS that usually assume straight motion of both the ego-vehicle and the opponent [2]. In the given example of Fig. 3, the motion vector of critical vehicle B intersects with the motion vector of our ego-vehicle E only in the very last moment.

Since today high-quality cameras come at a very low price, we aim at a vision-based solution. Furthermore, cameras are very flexible and easy to handle.

The framework overview of the developed system is shown in Fig. 4: With the help of the rectified image sequences a depth analysis is carried out. It is based on semi-global matching (SGM) using a real-time FPGA implementation [12], [15]. The detection of moving obstacles utilizes the so called Dynamic Stixel-World, a compact and efficient three-dimensional scene representation. This representation is introduced by Pfeiffer et al. in [19], [20] and is computed from the stereo-based depth information.

Based on this Stixel data, an object clustering is carried out. Up to this point, the developed approach is a generic detection and tracking scheme which can be used for any kind of traffic scene analysis.

Subsequently, geometrical assumptions are made to handle the special case at roundabout scenes. The object clustering results are used to compute a TTC that allows to decide whether a safe entrance into the roundabout is possible.

The article is organized as follows: Section II gives a detailed overview about the addressed challenges at roundabout scenarios and shows the results of our field of view analysis as well as the experimental camera set-up. The algorithms used to obtain the input data are sketched in Section III. This also includes a brief overview of the Stixel scene representation. Then, Section IV describes the clustering of Stixels to objects and the time-to-contact computation is

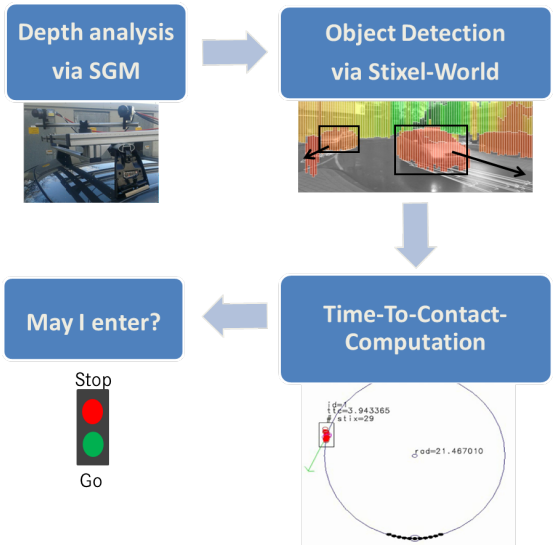


Fig. 4: The framework overview: Based on stereo depth analysis an object detection is carried out. Afterwards, for each object an expected collision time referred to the ego position is estimated. This time serves as an indicator in order to decide, if an entrance in the roundabout is safe or not.

presented in Section V. Results are given in Section VI: we show tracking results in different roundabouts and compare the decision of our situation analysis with the decision of test persons. As it turns out, our system mimics a careful and defensive driver. To handle turning vehicles, we use the results of the object-tracking procedure of Barth et al. [2] which estimates the yaw rate of oncoming vehicles. Section VII summarizes the article.

mu

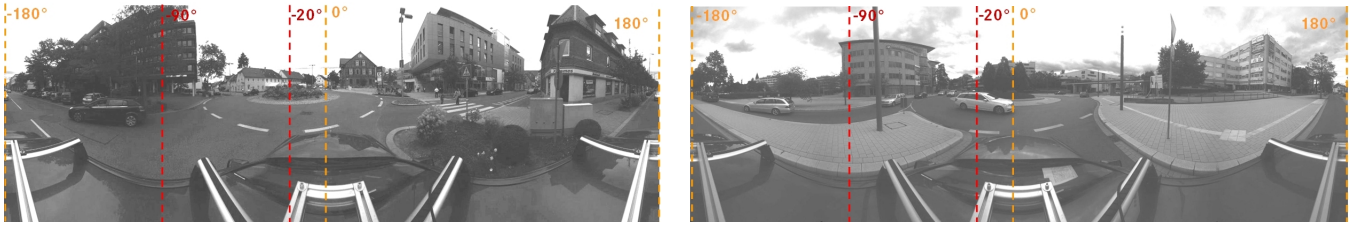


Fig. 5: Images of a spherical 360° camera showing typical urban roundabouts that are used for our field of view analysis. The painted vertical lines represent the viewing direction of the ego-vehicle. For the core application of tracking other vehicles in roundabouts, we focus on the field of view from -90° to -20° relative to the ego-vehicle (red lines). The orange lines represent the ego vehicle's viewing direction at $-180°$, $0°$ and $180°$. They are painted for a better image understanding.

II. The Challenges at Roundabout Scenarios and the Camera Set-up

A typical roundabout scenario with possible trajectories T_i of oncoming cars is sketched schematically in Fig. 3: while the ego-vehicle E is waiting for a risk-less entrance, vehicle B is driving inside the roundabout. At the same time, vehicle A is about to enter the circle as well. Furthermore Fig. 3 shows the entrance risk which is decoded as color from the ego vehicle's view. It is clear that the entrance risk grows with descending distance between the ego vehicle and an oncoming vehicle. Altogether, this is a challenging situation, because it allows to evolve in many very different, non-trivial ways.

For our task, we want to track oncoming traffic participants until they pass in front of us or make a right turn. It is evident that a forward looking camera set-up with a limited field of view is not sufficient for our task. This is because the two oncoming vehicles A and B (see Fig. 3) are not visible and therefore they would be detected too late. Thus, to achieve an optimal surveillance of the roundabout scene, we have to change the viewing direction and the field of view.

In a preliminary investigation we recorded traffic situations at about 50 different roundabouts in inner cities with the spherical camera system *Ladybug3*[3]. In most cases, one lane roundabouts with middle diameters between 15 m and 35 m were taking into account. Also, we recorded multi lane roundabouts with middle diameters up to 45 m.

With the help of the 360° images (see Fig. 5) the viewing directions of approaching vehicles referred to the ego vehicle heading direction were estimated. The results of our field of view analysis are summarized in a histogram which is shown in Fig. 6: about 95 % of all potentially relevant objects are located between -20° and -90° with respect to the horizontal viewing direction of the ego-vehicle.

Based on this data, we decided to work with 80° lenses looking at -50° to the left (see Fig. 7). The base line of our stereo camera system is 35 cm and the focal length is 837 px.

Next to the viewing direction we have to know which time another vehicle needs until it is in front of the ego-vehicle. This time is defined as the time-to-contact (TTC).

A short calculation reveals a traceable TTC: on typical inner city roundabouts with middle diameters of about 20 m people drive generally with 6-7 $\frac{m}{s}$. Accelerating up to this speed from a complete stop at the entrance takes about 2-3 s. Therefore, it is typically safe to enter the roundabout as long as no car is driving with that speed within the mentioned area.

[3]http://www.ptgrey.com/products/spherical.asp, viewed 30.09.2012

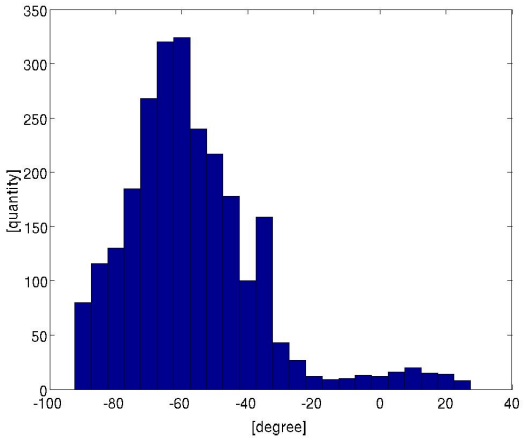
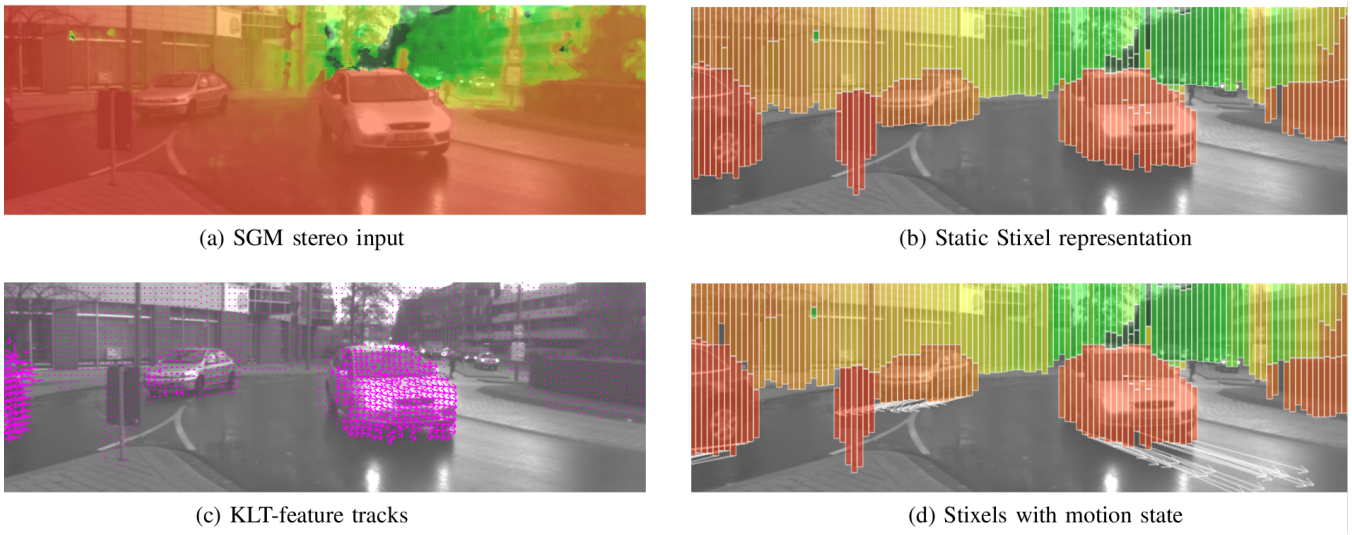


Fig. 6: Distribution of the viewing directions at approaching vehicles with respect to the heading direction of the ego-vehicle. It is shown that about 95 % of all observed vehicles move in the field of view from -20° to -90°.

III. 3 Dimensional Dynamic Scene Reconstruction via Stereo-Vision Methods

Detecting vehicles passing through roundabouts is achieved by using the rectified image sequences of our "left-side-looking" stereo camera system (see Section II).

At first, the dense disparity image is computed which is seen in Fig. 8(a). Due to its efficiency and performance, as shown in the Middlebury database [23], we use semi-global matching (SGM) [12], [15]. Gehrig et al. have introduced the first real-time implementation of SGM using a FPGA platform which estimates disparity images (up to 1400×400 px) with a rate of 25 Hz. The disparity range is between 0 and 127 px and a sub-pixel interpolation is carried out. For the further steps (object detection and tracking), the medium-level representation named Stixel-World proposed by Pfeiffer et al. [19] , [20] is computed from the disparity input data. A single Stixel is defined as a vertically oriented rectangle with a fixed width in the



(a) SGM stereo input

(b) Static Stixel representation

(c) KLT-feature tracks

(d) Stixels with motion state

Fig. 8: An exemplary roundabout is shown. The given images denote the different steps when computing Stixels for the current scenario. Fig. (a) shows the dense stereo input obtained from using SGM. The Color encodes the depth information: Red stands for very near objects, green for far away objects. Fig. (b) shows the corresponding static Stixel representation. Each Stixel has a 7 pixel width. The color encoding is the same as in Fig. (a). Fig. (c) shows the optical flow input used for tracking Stixels over time and Fig. (d) shows the final dynamic Stixel result with motion estimates. The arrows at the bottom points of each Stixel represent a 0.5 sec prediction.

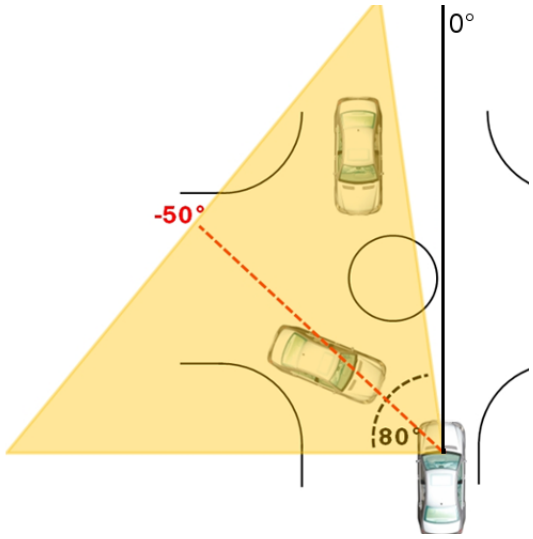


Fig. 7: The graphical representation of the chosen field of view (yellow rectangle) and the chosen viewing direction (red dotted line) of the new camera set-up. Assuming right-lane traffic, the camera stereo system is orientated to the left side.

image (e.g. 5 px) and a variable height. Every object within the image is approximated by a set of adjacent Stixels (see Fig. 8 (b)). This way, Stixels allow for an enormous reduction of the raw disparity input data, e.g. approximately $550,000$ disparity measurements from a 1400×400 px stereo image pair are reduced to a few hundred Stixels only. At the same time, Stixels give an easy access to the most task-relevant information such as freespace and obstacles and thus effectively bridge the gap between low-level (pixel-based) and high-level (object-based) vision.

According to [20], the three-dimensional scene is segmented into two different class types, namely ground and object. Both are expected as planar surfaces. The difference lies in their orientation: ground is expected as horizontal while object is assumed as vertical with a constant depth. Additionally, the segmentation is regularized by a set of physically motivated world model priors, such as gravity and ordering constraints. This way, the segmentation task leads to a typical maximum a posteriori (MAP) estimation problem. Solving for the most likely and thus optimum segmentation is achieved through the use of dynamic programming [3].

Up to this point, the Stixel representation only describes the current three-dimensional world geometry (in both the image and in 3D). However, deciding whether a roundabout is occupied by oncoming moving vehicles or not also requires additional velocity information.

For this purpose, the Stixel based tracking scheme proposed in [19] is chosen. Besides using stereo data, this scheme additionally requires optical flow information (see 8 (c)) as well as the ego vehicle's motion.

To this end, the first is computed by using the well-known feature-based KLT-tracker [25]. In [18], Pfeiffer compares different Optical Flow Estimators where the KLT-tracker performs best for this task. The ego vehicle's motion is extracted by using visual odometry [1] compensating the ego motion flow pattern. Here, it should be noted that this compensation is not necessary for a complete standing ego vehicle.

For estimating the motion properties of other objects, the obtained input data has to be combined properly. This is

achieved by following the 6D-Vision principle suggested by Franke et al. [10]. This scheme uses Kalman filtering [27] to estimate both the position and velocity of three-dimensional point feature. The result is combined in a rich 6-dimensional state vector. However, since in our considered scenarios all relevant objects are expected to move earthbound, this state vector allows to be reduced to 4D, namely the longitudinal $(Z,\dot{Z})$ and lateral $(X,\dot{X})$ state components referred to the left camera coordinate system, such that $\overline{X} = (X,Z,\dot{X},\dot{Z})^T$.

As a result, precise motion information is available for every Stixels independently. Stixels enriched with motion information are defined as dynamic Stixels. The Stixel-based tracking result for the exemplary scenario is depicted in Fig. 8(d). The described dynamic Stixel-World above has proven to be a powerful 3D representation which is also shown in [7], [6].

IV. Clustering process

The goal of the following steps is to estimate the center position and the mean velocity of relevant vehicles in roundabouts which allows an efficient estimation of the TTC. For this purpose, the independent Stixels $s_i \in \{1,..,\mathrm{I}\}$ have to be grouped to only a small number of relevant objects c_k with $k \in \{1,..,\mathrm{K}\}$ and $\mathrm{K} << \mathrm{I}$. A successful clustering is based on well-considered geometrical and physical conditions. To this end, the following assumptions are made:

- **minimum number of Stixels**: due to their horizontal expansion, objects are represented by a minimum number of Stixels.
- **geometrical characteristics**: the euclidean distance between two Stixels s_i and s_j is a relevant criterion for the spatial separation.
- **physical characteristics**: Stixels representing the same object have a uniform velocity and driving direction.

We applied a real-time clustering procedure which is based on the often cited and common DBSCAN algorithm [9].

The key idea is that an arbitrary Stixel s_i of a cluster c_k has at least a minimum number of Stixel neighbors within a given neighborhood threshold ε. In our approach, we assume that the number of Stixels within a cluster is considerably higher than outside of a cluster. Thus, clusters which contains only a very small number of Stixels are flagged as noise.

Even though the spatial Euclidean distance d is frequently used as a neighborhood criterion, the DBSCAN [9] allows any kind of cost functions. For our application, it is not sufficient to use the spatial Euclidean distance $\mathrm{d} = \mathrm{dis}(s_i,s_j)$ as the only neighborhood constraint. This is because side-by-side driving cars with different driving directions tend to be merged to one object. For an example, in a two lane roundabout a turning and a passing vehicle (which are very close to each other) are fused together.

To this end, a second neighborhood criterion is defined which is the angle ϕ between the two motion vectors $u_i = (\dot{X},\dot{Z})_i^T$ and $u_j = (\dot{X},\dot{Z})_j^T$ of the Stixels s_i and s_j:

$$\phi = arccos\left(\frac{< u_i, u_j >}{\|u_i\| \ \|u_j\|}\right) \tag{1}$$

A drawback of the standard DBSCAN algorithm is its quadratic complexity $O(n^2)$, where in our case n equals the number of Stixels I. To reduce the resulting computational burden, we use the modified l-DBSCAN [26] which is a hybrid clustering method with a runtime complexity of $O(n)$. Its key idea is to start with a coarser clustering of the complete data set, where each cluster is represented by its leader point l. Then, a fine clustering is carried out for which only those leader points are considered. Finally, after the grouping process, each cluster is represented by its center position X_k, Z_k and its mean velocity $\dot{X}$ and $\dot{Z}$.

V. The time-to-contact-computation

In the following, the estimation of the TTC_k is extracted which is performed at each time-step and for each detected cluster k. The goal is to predict whether a safe entrance into the roundabout is possible or not. With the help of the center points of each cluster and a nearest neighbor criterion, other vehicles driven trajectories are estimated (see Fig. 10).

Due to the described challenges in Section II the assumption is made that detected vehicles are a potential threat for the ego vehicle. Consequently, it is supposed that each oncoming vehicle will pass the ego vehicle and that the motion trajectory within the roundabout is approximately represented by a circular shape. This statement may not hold for all oncoming vehicles (e.g. vehicles which will leave the roundabout before the ego vehicle's position , see Fig. 1) but from the developer's viewpoint it is a safe assumption.

To estimate the TTC_k, a circle is fitted to the driven mean positions $[X_k,Z_k]_t$ as soon as a cluster c_{kt} is steadily observed over time $t \in \{1,...,\mathrm{T}\}$. The method of [5] is used for the circle estimation which is based on a least square fit. Hereby, the sum of the squares

$$F = \sum_{t=1}^{T} {v_t}^2 \tag{2}$$

is minimized where v_t is the error distance function defined as:

$$v_t = \sqrt{(X_{kt}-a_k)^2 \ + \ (Z_{kt}-b_k)^2} - R_k \ , \tag{3}$$

with the circle radius R_k and the circle center $[a,b]_k$. At the beginning of the tracking only a few object points $[X,Z]_{kt}$ are available and its accuracy is limited. As a result, the quality of the estimated circle radius is quite poor. To overcome this problem and to allow a robust and reliable estimation at any time, the possible size of the circle radius R_k is constrained. Therefore, it is assumed that digital maps (e.g. navigation systems) will provide a geometrical information.

Furthermore, the length of the circular arc ca_k which a vehicle will drive up to a possible collision point is defined by the circle angle α. This angle α is calculated from the current vehicle position, the circle center and the position of the ego-vehicle (see Fig 9). The estimation of ca_k is straightforward:

$$\mathrm{ca}_k = \pi \ R_k \ \frac{\alpha}{180^\circ} \ . \tag{4}$$

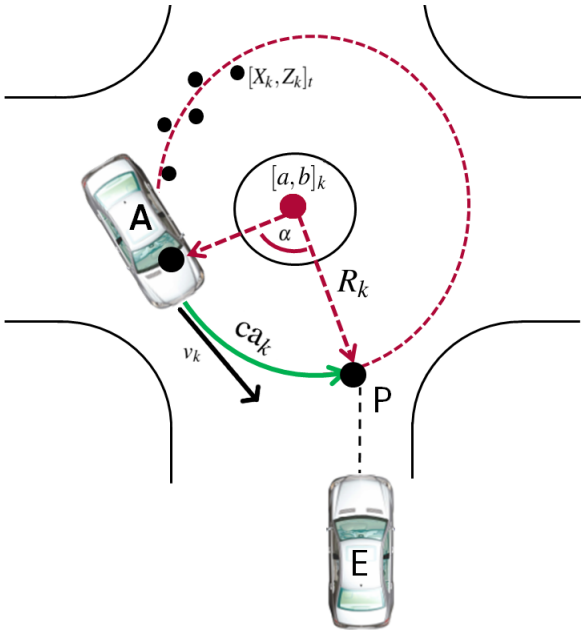


Fig. 9: The mentioned geometrical parameters of the TTC-computation: With the help of the driven mean positions $[X_k, Z_k]_t$ and a possible collision point P a robust circle fit is computed which delivers the circle radius R_k and the circle center $[a,b]_k$. With the use of the current velocity v_k of object A an expected TTC-computation is estimated.

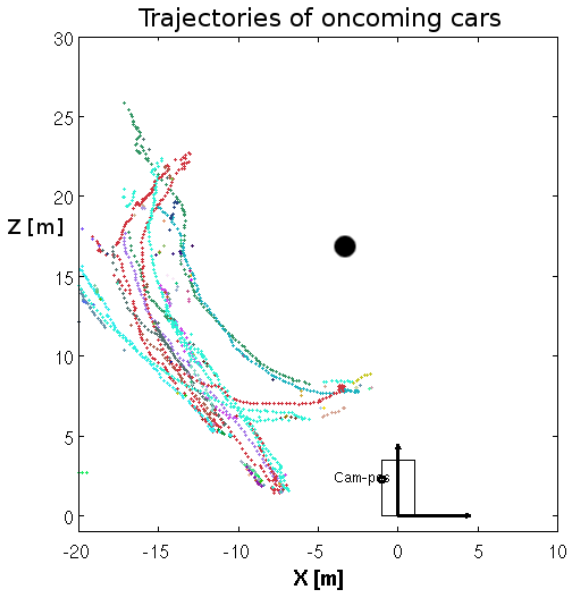


Fig. 10: The estimated trajectories of our tracking algorithm for a sample 80 s sequence at a two lane roundabout. The color represents the different vehicle tracks. Roughly, 50 percent of all oncoming vehicles drove a circular arc and passed the ego-vehicle (black rectangle). All other vehicles were leaving the roundabout.

Finally, the TTC_k is determined by the estimated velocity $v_k = \sqrt{(\dot{X}_k^2 + \dot{Z}_k^2)}$ and ca_k:

$$\mathrm{TTC}_k = \frac{\mathrm{ca}_k}{v_k} \ . \tag{5}$$

If the TTC_k is below a given time threshold, the system advises not to enter the roundabout. The TTC_k is updated at each time step which is exemplary shown in Fig. 11.

VI. Results

For our experiments, we have evaluated video material of typical roundabout scenarios recorded at rush-hour traffic. The stereo camera set-up has the configuration as described in Section II. The images are cropped to 1400×400 px to focus on the relevant scene content. The stereo computation using the SGM is running on an FPGA (Xilinx Virtex-4) within 40 ms which produces one frame delay. The dynamic Stixel algorithm, the clustering process and the TTC computation is running on the CPU (Intel Core i7-980X 3.33Ghz) in real-time within ∼40 ms. The images are captured with 25 Hz.

A. Results of the vehicle tracking and the TTC computation

Fig. 10 shows different tracking samples of a 80 s sequence of the two lane roundabout from Fig. 8. The oncoming cars are recognized at an approximate distance of 25 m. It is visible from the processed data that vehicles which passed us drove a circular arc.

Due to side-by-side driving, in some cases, tracks of covered vehicles were lost and a new cluster re-initialization had to occur. Scattered outliers are observed at distances of approximately 20-25 m which, however, has shown no negative influence on the TTC computation.

Fig. 11 shows the dynamic Stixel representation with the TTC computation of a typical roundabout scene where an oncoming car passed the ego-vehicle. The color of the Stixels represents the distance with respect to the ego-vehicle position. Red stands for near and green for objects far away. The drawn arrows describe the direction the Stixels will move in the next 0.5 s.

At this time the system is not able to decide, if an oncoming vehicle will turn or will pass the ego vehicle because only the assumption is made that every vehicle will pass the ego position. In Section VI-C a solution is described how to separate between turning and passing vehicles.

B. Evaluation with test persons

In order to evaluate the estimated TTC output of the system, the estimated stop and go phases have been compared with the driving behavior of test persons. They were shown recorded scenes of one lane roundabouts and had to mark those time windows where they would enter the roundabout.

Fig. 12 compares the results of each test person with our algorithm using a 90 s sequence with 19 different traffic situations. Note that this is just an extract of our evaluation with about 50 independent roundabout scenarios.

Generally, the red-green-phases of the algorithm coincide with the phases of the test persons. Thereby, the geometrical

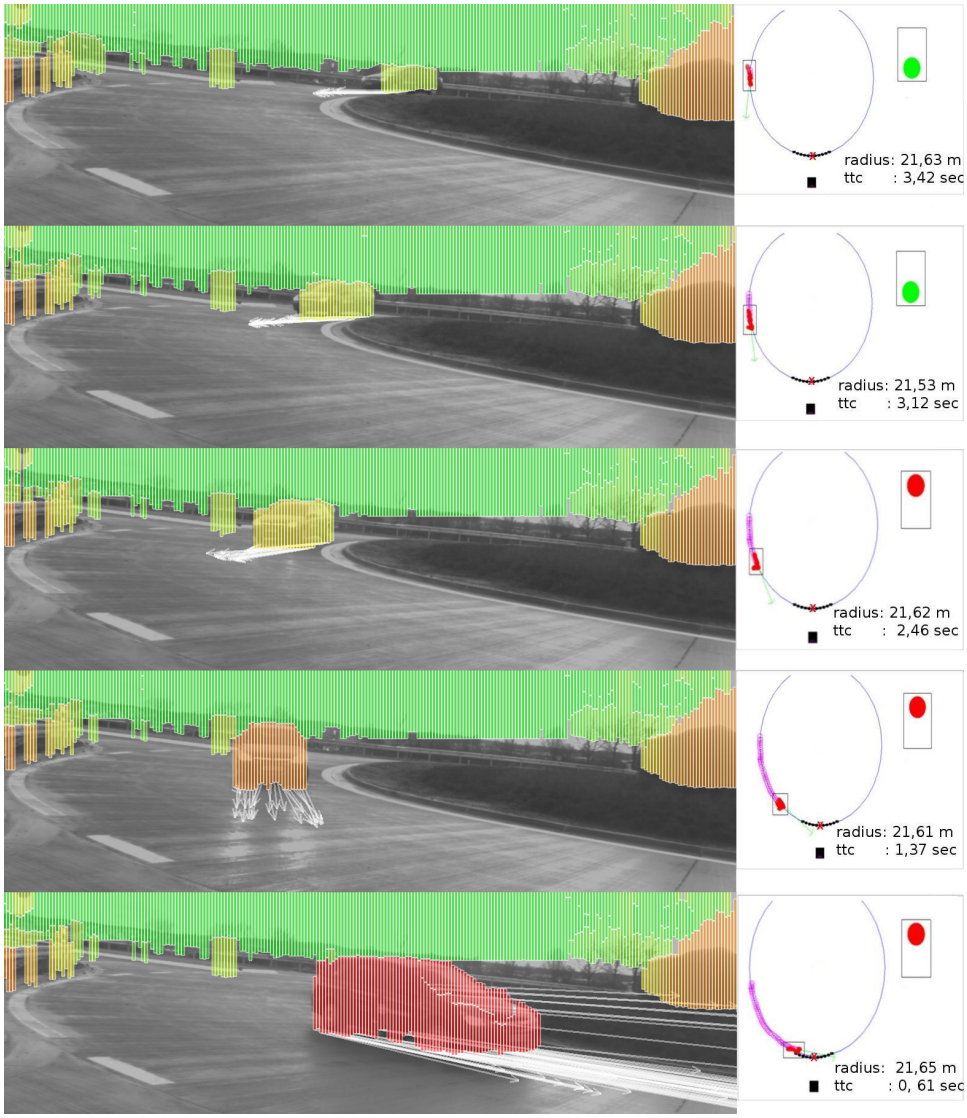


Fig. 11: A sequence set of a typical scene with an oncoming vehicle at a roundabout. The images are illustrated together with the corresponding dynamic Stixel representation. On the right side the results of the clustering process and the TTC computation is shown for each scene.

assumptions of the driving behavior and the TTC threshold of 2.5 s are confirmed. Scene 2 (about 0:05 s) shows how a traffic situation could be interpreted differently which underlines the general complexity at roundabout scenarios. In this scene, our Stop-and-Go Assistance System does not allow to enter the roundabout which is a good example for the defensiveness of our decision strategy.

However, this evaluation also reveals certain limits of our approach: In contrast to the test persons, the TTC computation can not "recognize" turning vehicles, such as shown in Scene 3 (about 0:18 s), Scene 9 (about 0:50 s) and Scene 15 (about 0:70 s). For example, in Figure 13 the results of a turning vehicle are presented in detail: Due to our made assumption that all oncoming cars will pass the ego vehicle, the approach does not allow an entering at the end of the scene, even though an entrance is principally possible.

In the last section of this article, a solution for the detection of turning vehicles is presented which is based on a full motion state estimation of oncoming vehicles.

C. Detection of turning vehicles using yaw-rate estimation

In our case it is not sufficient enough to use only the information about position and velocity of current objects to predict if the vehicle will turn or not. Next to this information, particularly the yaw-rate is another important motion state parameter which should be considered.

At a roundabout scenario the following assumptions for oncoming cars are conclusive: A vehicle which drives continuously on a circular path should have a constant yaw-rate. In contrast to this behavior, the driving path of a turning vehicle could be described by a transition curve or more strictly as a clothoid. Here, the yaw-rate is continuously decreasing until it is zero at the turning point of the curve. Subsequently, this state parameter is increasing in the opposite direction.

For the estimation of the yaw-rate we rely on the vehicle tracking approach as proposed by Barth et al. [2]. In contrast to our Stixel-based approach, this idea uses a set of tracked 3D feature points on the object's surface and a vehicle motion model to extract the pose and motion state of a vehicle, including acceleration and yaw rate. An extended Kalman filter is used to estimate the object state in a time-recursive fashion given measurements of tracked feature point positions on the image plane and stereo disparities derived from stereo vision.

Figure 14 shows the results of this approach for a turning and a passing vehicle. Next to the image sequence the figure presents the behavior of the velocity and the yaw-rate during the tracking. At first, in both cases the yaw-rate increases due to the initialization of the Kalman filter. After 1.5 s the passing vehicle has a nearly constant positive yaw-rate which means that the car drives a left curve referred to the ego position. As expected, the turning vehicle's yaw-rate shows a complete different characteristic: After 1.0 s the yaw-rate is nearly zero. The switch to a negative yaw-rate indicates exactly the change from a left curve to a right curve which is a typical behavior for the most turning vehicles. This "yaw-rate-flip" is a strong indicator, that an oncoming car will leave the roundabout.

VII. SUMMARY

In this article, a stereo-based time-to-contact computation for right of way situations at roundabouts was presented. For this purpose, urban roundabouts were observed with a spherical 360° camera system to configure an optimal stereo camera set-up.

Dense disparity images are used to compute the dynamic Stixel-World which is a compact three-dimensional environment representation for urban traffic situations. This work proves the power of the dynamic Stixels which support our processing steps.

A well known clustering method was used to group independent dynamic Stixels representing the same object. This procedure allows reliable tracking of oncoming vehicles at urban roundabouts. In order to handle such complex situations properly, we assume that all tracked vehicles drive on a circular arc. This has proven a defensive but safe assumption.

For a reliable time-to-contact computation, a robust circle fit method was used which is supported by additional geometric constraints.

The system's estimated stop and go phases have been compared to the driving behavior of 10 different test persons. According to these tests, it has performed no misjudgment of the current right-of-way situations.

For the detection of turning vehicles it is not sufficient enough to estimate only the the center of mass and the velocity of relevant objects. In our case we presented a 3D point tracking algorithm which estimates the complete motion state of oncoming vehicles. It turns out that the estimated yaw-rate is a reliable indicator if an vehicle will turn next to the ego vehicle in a roundabout scenario.

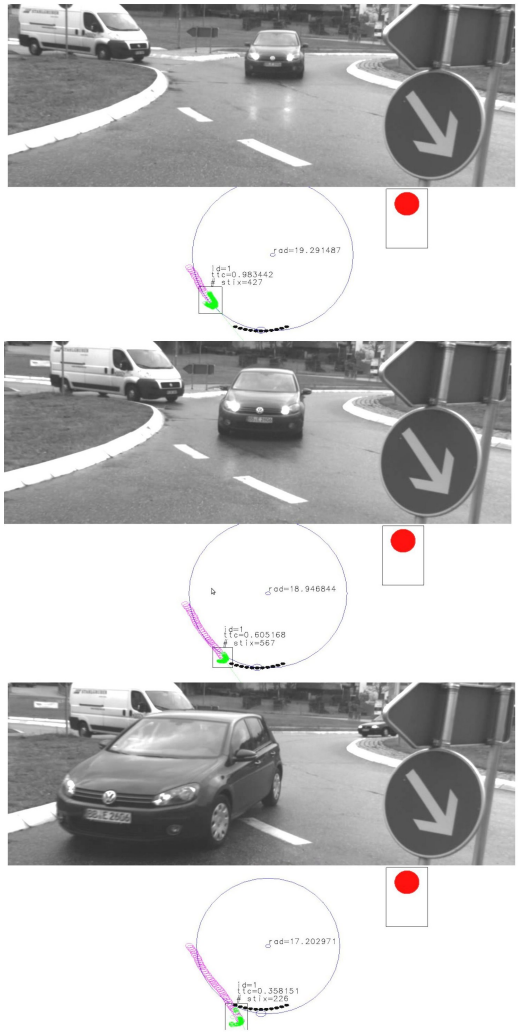


Fig. 13: An example scene of a turning vehicle: Due to the geometrical assumptions the approach does not allow to enter the roundabout, although an entrance is principally possible.

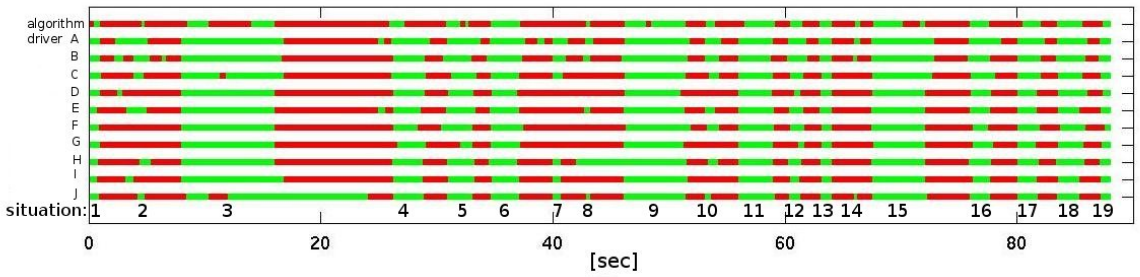


Fig. 12: The estimated stop and go phases (red and green) of our algorithm compared to the behavior of test persons for a 90 s sequence of a typical urban roundabout. The first line represents the "decision" of our algorithm. The next 10 lines represent the virtual driving behavior of our test persons A-J. In 13 of 19 independent traffic situations (1, 4-6, 9, 11-13, and 15-19) the algorithm decision closely corresponded to the human behavior. In two cases (situation 8 and 14), the algorithm only matched to five or less participants. Again, in four situations (3, 5, 9, and 15) the algorithm switched to red for a few frames while some of the participants decided otherwise. In these cases, the test persons recognized that the vehicles turned.

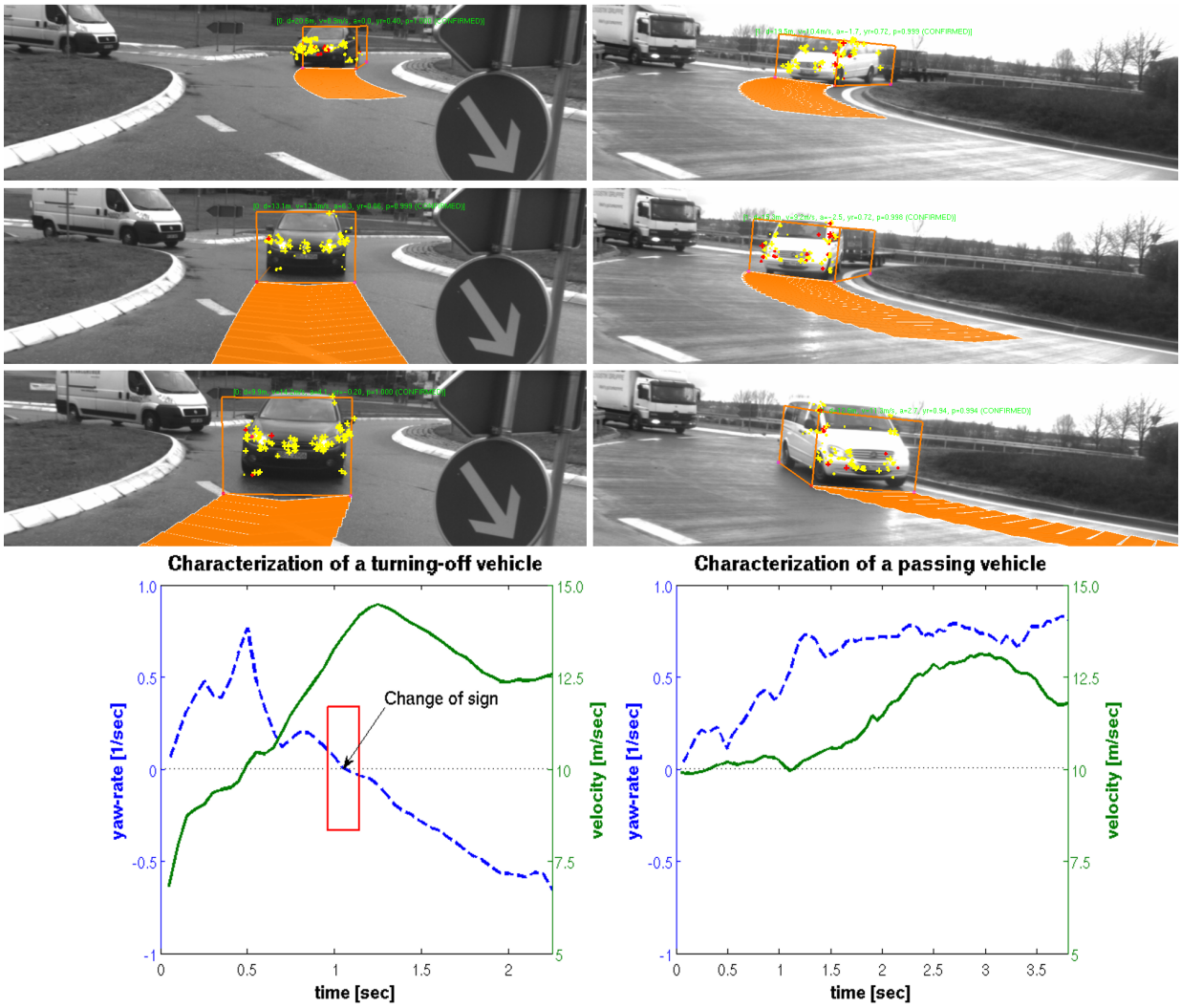


Fig. 14: The results of the vehicle tracking procedure proposed by Barth et al. [2] for a turning vehicle (left) and a passing vehicle (right). The yellow and red points on the objects are tracked 3D feature points. The orange bounding box represents the spatial extension of the object and the orange area is the predicted driving corridor for the next 0.5 s. As expected, the yaw-rate behavior between both cars is very different. The change of sign of the yaw-rate is a strong indicator, that an oncoming car leaves the roundabout.

VIII. About the Authors

Maximilian Muffert studied geodesy and geoinformation at the University of Bonn, Department of Photogrammetry. The supervisor for his master thesis was Prof. Dr.-Ing. Dr. h.c. mult. Wolfgang Förstner. Today, Maximilian is working towards his Ph.D. in Photogrammetry. For this purpose, he is employed at the Daimler AG in Sindelfingen, Germany. He is a member of the team GR/FFE lead by Dr. Uwe Franke. His area of research is the modeling of inner city environments with extraordinary stereo camera set-ups. His supervisor at the University of Bonn remains Wolfgang Förstner.

David Pfeiffer studied computer science and received his diploma at the Humboldt-University of Berlin in 2008. Since 2008 David is employed at the Daimler AG in Sindelfingen, Germany where he is a member of team GR/FFE lead by Uwe Franke. In 2011 he received the Ph.D. degree in computer science for his work on the modeling of dynamic 3D environments. His supervisor was Prof. Dr. rer. nat. Ralf Reulke, Humboldt-University of Berlin. His further areas of research is soft- and hardware optimization for real-time stereo vision.

Uwe Franke received the Ph.D. degree in electrical engineering from the Technical University of Aachen, Germany, in 1988 for his work on content based image coding. Since 1989 he has been with Daimler Research and Development and has been constantly working on the development of vision based driver assistance systems. He developed Daimler's lane departure warning system ("Spurassistent", available since 2000). Since 2000 he has been head of Daimler's Image Understanding Group and is a well known expert in real-time stereo vision and optical flow. Recent work is on optimal fusion of stereo and motion, called 6D-Vision and scene flow. In 2002, he was program chair of the IEEE Intelligent Vehicles Conference in Versailles, France.

References

[1] Hernán Badino. A robust approach for ego-motion estimation using a mobile stereo platform. In 1st *International Workshop on Complex Motion, IWCM*, Günzburg, Germany, October 2004. Springer.

[2] Alexander Barth and Uwe Franke. Tracking oncoming and turning vehicles at intersections. In *IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 861–868, Madeira Island, Portugal, September 2010.

[3] Richard Bellman. *Dynamic Programming*. Princeton University Press, 1957.

[4] Adam Berthelot, Andreas Tamke, Thao Dang, and Gabi Breuel. Stochastic situation assessment in advanced driver assistance system for complex multi-objects traffic situation. In *IEEE/RSJ International Conference on Intelligent Robots and Systems(IROS)*, pages 1180–1185, Vilamoura-Algarve, Portugal, December 2012.

[5] Nikolai Chernov. *Circular and linear regression: Fitting circles and lines by least squares*. CRC Press, Inc., Boca Raton, FL, USA, 2010.

[6] Markus Enzweiler, Matthias Hummel, David Pfeiffer, and Uwe Franke. Efficient stixel-based object recognition. In *IEEE Intelligent Vehicles Symposium (IV)*, pages 1066–1071, Alcala de Henares, Spain, June 2012.

[7] Friedrich Erbs, Beate Schwarz, and Uwe Franke. Stixmentation - probabilistic stixel based traffic scene labeling. In *British Machine Vision Conference (BMVC)*, pages 71.1–71.12, Surrey, England, September 2012.

[8] Andreas Ess, Bastian Leibe, Konrad Schindler, and Luc J. Van Gool. Moving obstacle detection in highly dynamic scenes. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 56–63, 2009.

[9] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *KDD*, pages 226–231, 1996.

[10] Uwe Franke, Clemens Rabe, Hernán Badino, and Stefan Gehrig. 6d-vision: Fusion of stereo and motion for robust environment perception. In *German Association for Pattern Recognition (DAGM)*, Vienna, Austria, September 2005.

[11] Behörde für Stadtentwicklung und Umwelt. Planungshinweise für Stadtstrassen: Knotenpunkte & Kreisverkehre, Freie und Hansestadt Hamburg, Germany, 2009.

[12] Stefan Gehrig, Felix Eberli, and Thomas Meyer. A real-time low-power stereo vision engine using semi-global matching. In *International Conference on Computer Vision Systems (ICVS)*, 2009.

[13] Christoph Hermes, Julian Einhaus, Markus Hahn, Christian Wöhler, and Franz Kummert. Vehicle tracking and motion prediction in complex urban scenarios. In *IEEE Intelligent Vehicles Symposium (IV)*, pages 26–33, San Diego, CA, USA, June 2010.

[14] Michael Himmelsbach, Andre Mueller, Thorsten Luettel, and Hans-Joachim Wuensche. Lidar-based 3d object perception. In *Proceedings of 1st International Workshop on Cognition for Technical Systems*, Munich, Germany, October 2008.

[15] Heiko Hirschmüller. Accurate and efficient stereo processing by semi-global matching and mutual information. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 807–814, San Diego, CA, USA, June 2005.

[16] Dominik Kellner, Jens Klappstein, and Klaus Dietmayer. A grid-based dbscan for clustering extended objects in radar data. In *IEEE Intelligent Vehicles Symposium (IV)*, pages 365–370, Alcala de Henares, Spain, June 2012.

[17] Maximilian Muffert, Timo Milbich, David Pfeiffer, and Uwe Franke. May i enter the roundabout? A time-to-contact computation based on stereo-vision. In *IEEE Intelligent Vehicles Symposium (IV)*, pages 565–570, Alcala de Henares, Spain, June 2012.

[18] David Pfeiffer. *The Stixel World*. PhD thesis, 2012.

[19] David Pfeiffer and Uwe Franke. Efficient representation of traffic scenes by means of dynamic Stixels. In *IEEE Intelligent Vehicles Symposium (IV)*, pages 217–224, San Diego, CA, USA, June 2010.

[20] David Pfeiffer and Uwe Franke. Towards a global optimal multi-layer Stixel representation of dense 3D data. In *British Machine Vision Conference (BMVC)*, Dundee, Scotland, August 2011. BMVA Press.

[21] Sylvia Pietzsch, Nils Appenrodt, Juergen Dickmann, and Bernd Radig. Model-based fusion of laser scanner and radar data for target tracking. In *8th International Workshop on Intelligent Transportation*, pages 44–49, Hamburg, Germany, March 2011.

[22] Nicolas Saunier, Tarek Sayed, and Clark Lim. Probabilistic collision prediction for vision-based automated road safety analysis. In *The 10 th International IEEE Conference on Intelligent Transportation Systems*, pages 872–878, Seattle, USA, October 2007.

[23] Daniel Scharstein and Richard Szeliski. Middlebury online stereo evaluation, 2002. http://vision.middlebury.edu/stereo.

[24] Sebastian Thrun. What we're driving at. `http://googleblog.blogspot.de/2010/10/what-were-driving-at.html`, October 2010.

[25] Carlo Tomasi and Takeo Kanade. Detection and tracking of point features. Technical report, School of Computer Science, Carnegie Mellon University, April 1991.

[26] P. Viswanath and Rajwala Pinkesh. l-dbscan : A fast hybrid density based clustering method. In *ICPR*, pages 912–915, 2006.

[27] Greg Welch and Gary Bishop. An introduction to the kalman filter. Technical report, Department of Computer Science, University of North Carolina at Chapel Hill, 1995.